ANCHOR: a web server and database for analysis of protein–protein interaction binding pockets for drug discovery

Lidio M. C. Meireles[1], Alexander S. Dömling[2,3,*] and Carlos J. Camacho[1]

[1]Department of Computational Biology, [2]Department of Pharmaceutical Sciences and [3]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15261, USA

ABSTRACT

ANCHOR is a web-based tool whose aim is to facilitate the analysis of protein–protein interfaces with regard to its suitability for small molecule drug design. To this end, ANCHOR exploits the so-called anchor residues, i.e. amino acid side-chains deeply buried at protein–protein interfaces, to indicate possible druggable pockets to be targeted by small molecules. For a given protein–protein complex submitted by the user, ANCHOR calculates the change in solvent accessible surface area (ΔSASA) upon binding for each side-chain, along with an estimate of its contribution to the binding free energy. A Jmol-based tool allows the user to interactively visualize selected anchor residues in their pockets as well as the stereochemical properties of the surrounding region such as hydrogen bonding. ANCHOR includes a Protein Data Bank (PDB) wide database of pre-computed anchor residues from more than 30 000 PDB entries with at least two protein chains. The user can query according to amino acids, buried area (SASA), energy or keywords related to indication areas, e.g. oncogene or diabetes. This database provides a resource to rapidly assess protein–protein interactions for the suitability of small molecules or fragments with bioisostere anchor analogues as possible compounds for pharmaceutical intervention. ANCHOR web server and database are freely available at http://structure.pitt.edu/anchor.

INTRODUCTION

Protein–protein interactions (PPIs) are attractive targets for pharmaceutical intervention (1–5) because their ubiquitous role in mediating biological processes in the cell and the fact that many diseases such as cancer can be attributed to malfunctioning PPIs (6–8). The ability to modulate specific PPIs with small organic molecules for therapeutic applications has therefore been pursued by the scientific community, who faces the challenging task of discovering and/or designing small molecules that bind with high affinity to relatively large and flat protein–protein interfaces.

Despite the fact that proteins often interact through large contact surfaces, the presence of well-defined 'anchor' sites and cavities which when filled with the appropriate compound might trigger a strong attraction between receptor and ligand (9,10) allows medicinal chemists to focus on targeting these areas. Alanine scanning mutagenesis has been extensively used to detect the amino acid residues that contribute to the binding free energy of a given PPI (11–13). In addition, a large number of computational methods have been developed to predict 'hotspots', i.e. those residues that result in significant loss of binding affinity when mutated to alanine ($\Delta\Delta G > 2.0$ kcal/mol) (14–18), making use of the wealth of experimental data available from alanine substitution studies to train their models. However, few studies have focused exclusively on anchor sites (10), which contrary to hotspots have an explicit concave/convex geometry appealing for pharmaceutical intervention.

The identification of anchor residues in PPIs is very useful not only to provide insights into mechanisms of protein–protein recognition, but also to indicate the areas to be targeted with small molecules. Here, we report the development of ANCHOR, a web-based tool created to facilitate the analysis of PPI druggable cavities. For a given protein–protein complex submitted by the user, ANCHOR calculates the change in solvent accessible surface area (ΔSASA) upon binding for each side-chain, along with an estimate of its contribution to the binding free energy (19,20). A Jmol-based tool allows the user to interactively visualize selected anchor residues in their

pockets as well as the stereochemical properties of the surrounding region such as hydrogen bonding and charge–charge interactions. Moreover, ANCHOR includes a database of pre-computed anchor residues from more than 30 000 Protein Data Bank (PDB) (21) entries with multiple protein chains. The user can query the database according to amino acids, buried area (ΔSASA), energy or keywords related to indication areas, e.g. oncogene or diabetes. These queries could be useful to rapidly screen for suitable sites/cavities that fit fragments with chemical properties similar to anchor residues, correlating targets with functional categories or diseases. ANCHOR is complementary to existing tools for interface analysis of proteins reviewed recently (22).

MATERIALS AND METHODS

Characterization of anchor residues

For a given protein–protein complex structure, ANCHOR performs the following calculations:

(i) Add missing atoms including polar hydrogen using CHARMM19 (23) and perform a small round of hydrogen minimization to optimize hydrogen bonding.

(ii) Calculate the change in solvent-accessible surface area upon binding for each residue's side-chain i (ΔSASA$_i$; C-alpha included) by computing the difference in SASA of the side-chain in the unbound protein (isolated from complex) and in the bound protein complex: $\Delta SASA_i = SASA_i^{\text{unbound}} - SASA_i^{\text{bound}}$. SASA$_i$ is calculated with NACCESS (Hubbard and Thornton, 'NACCESS', Computer Program, 1993) using default parameters and a 1.4 Å water radius. The user can specify the chains that comprise the receptor and ligand or choose the default settings, in which case the value of ΔSASA$_i$ is obtained for every residue of each individual protein chain (unbound) against all the others (bound).

(iii) Estimate the associated binding free energy of each residue using FastContact (19,20), a fast empirical pairwise estimate that combines a standard distance-dependent dielectric '4r' electrostatic and a desolvation contact potential (24). FastContact has been successfully applied in protein-protein docking (25,26) and for scoring different sets of docked conformations (20).

(iv) Output is the values of ΔSASA and FastContact energy for each residue.

Database and query engine

We applied the procedure described above to construct a database of pre-computed anchor residues from 30 737 PDB entries with at least two protein chains (but no DNA/RNA chains). For NMR structures, we used the first NMR model deposited in the PDB as a representative structure of the NMR ensemble. Since anchor residues are constrained on the protein–protein interface, the results

do not change significantly by using a different model from the NMR ensemble. For X-ray structures with resolution better than 4.0 Å, we computed the anchor residues from the most probable biological assembly predicted by the European Bioinformatics Institute PISA service (27), which has been reported to predict biological assemblies with high accuracy (28). By using biological assemblies, we intended to enrich the database with anchor residues from biologically meaningful protein interfaces, as opposed to crystallographic interfaces.

Only the residues bearing a minimal change of SASA (ΔSASA > 0.5 Å^2) or whose binding energy was estimated to be less than -0.5 kcal/mol were included on the database. All other residues were discarded. We intentionally set such low thresholds for ΔSASA and energy in order to allow maximum user flexibility in querying the database, while avoiding overloading it with millions of non-interacting residues. On the construction of the database, we used the software's default behavior of treating each individual chain in a PDB file as an individual protein, thus the values of ΔSASA and energy recorded on the database correspond to the interaction of each individual chain against all the others.

The ANCHOR database is stored and indexed by MySQL (http://www.mysql.com), a relational database system. The user may query for anchor residues in PPIs by attributes such as PDB ID, residue type, ΔSASA, energy and/or keywords from the PDB file (Figure 1). To process keyword queries, ANCHOR uses a web service API (Application Programming Interface) to first obtain online from the RCSB PDB servers (http://www.pdb.org) a list of PDB entries satisfying the given keyword query. Then, without any user intervention, ANCHOR returns all anchor residues from those PDB entries and satisfying the remaining attributes of the query (if applicable). Because keyword queries are in fact processed on-demand by RCSB PDB, ANCHOR inherits all its complex keyword query features, such as support for logical operators, grouping terms, wildcard, etc. For example, the following is a valid keyword query accepted by ANCHOR (taken from the RCSB PDB

Figure 1. ANCHOR database. Front page of the search engine.

documentation): *BCL AND (apoptosis OR 'programmed cell death')*.

An example of the output of a query is shown in Figure 2. It consists of a list of PDB codes, chain ids, residue name and number, the amount of solvent accessible surface area buried by the side chain (ΔSASA) both in Å^2 and as percentage relative to the SASA of the side-chain in the free tri-peptide Ala-Res-Ala. The list also includes an empirical estimate of the residue contribution to the binding free energy, keywords from the PDB, and PDB title.

Visualization tool

The visualization tool is based on Jmol (http://www.jmol .org) and needs the Java Virtual Machine (JVM) to be pre-installed in the host computer (Figure 3). The interactive tool appears as a new window that includes visualization options as well as a table with all the interface residues of the selected chain sorted according to ΔSASA. Any of the residues listed in the table can be selected in order to visually inspect its chemical complementarity on the acceptor protein. The anchor is displayed as sticks (default), and the acceptor protein as a solvent accessible surface (default). Surface options such as molecular surface or cavities, color codes based on partial charges (default) or B-factors, transparency and environment are available to the display mode. Other options such as atom representation, display, color scheme, and labels are applied to a selection before placing on view ('apply to' option). The user can further display other residues from the table by clicking the appropriate

residue number, or select the environment option to expand the view of the anchor to residues within 6, 8, 10 and 12 Å or the whole protein. Other settings include: hydrogen bonds shown with a dashed line, zooming in around a residue, and secondary structures shown in cartoon representation.

UTILITY AND DISCUSSION

Analysis of PPI hotspots for drug discovery

In order for a small molecule to bind with high affinity to a given protein target, the molecule ideally needs to bury a substantial fraction of its surface area into the protein surface. For enzymes and small molecule receptors, this is typically accomplished by targeting the large cavities present on the surface of those proteins (4). However, for PPI targets, which are often devoid of such large cavities, the burial of large amounts of surface area can be achieved by simultaneously targeting the sites of two or more anchor residues lying in close proximity to each other. This suggestion is corroborated by the observation made by Fuller and colleagues (29) that PPIs and PPI inhibitors tend to occupy simultaneously several average-sized pockets. Intuitively, the larger the number of anchor residues and their associated ΔSASA more 'druggable' is the protein interface with a natural molecular weight limit for small molecular weight inhibitors.

As an example, Figure 3 (anchor visualization tool) shows the interaction between MDM2 (surface representation) and three anchor residues from p53 (stick representation). These three residues are among those with the

PDB	Ch.	Res.	Res.#	ΔSASA (Å^2)	ΔSASA (%)	Energy (kcal/mol)	Keywords	PDB Title
2W96	A	TRP	150	151.4	71.7	-6.7	CELL CYCLE	CRYSTAL STRUCTURE OF CDK4 IN COMPLEX WITH A D-TYPE CYCLIN
2W9Z	A	TRP	150	151.1	71.5	-6.7	TRANSFERASE	CRYSTAL STRUCTURE OF CDK4 IN COMPLEX WITH A D-TYPE CYCLIN
3G33	B	TRP	150	149.5	70.8	-6.0	CELL CYCLE	Crystal structure of CDK4/cyclin D3
3G33	F	TRP	150	149.5	70.8	-6.0	CELL CYCLE	Crystal structure of CDK4/cyclin D3
2W9F	A	TRP	150	147.6	69.8	-6.7	CELL CYCLE	CRYSTAL STRUCTURE OF CDK4 IN COMPLEX WITH A D-TYPE CYCLIN
3G33	H	TRP	150	146.6	69.4	-6.5	CELL CYCLE	Crystal structure of CDK4/cyclin D3
3G33	D	TRP	150	146.6	69.4	-6.5	CELL CYCLE	Crystal structure of CDK4/cyclin D3
2W99	A	TRP	150	144.7	68.5	-6.9	CELL CYCLE	CRYSTAL STRUCTURE OF CDK4 IN COMPLEX WITH A D-TYPE CYCLIN
3EQS	E	TRP	7	142.4	67.4	-5.7	Ligase	Crystal structure of human MDM2 in complex with a 12-mer peptide inhibitor
3EQS	B	TRP	7	142.4	67.4	-5.7	Ligase	Crystal structure of human MDM2 in complex with a 12-mer peptide inhibitor
2Z5S	B	TRP	23	139.0	65.9	-5.9	CELL CYCLE	Molecular basis for the inhibition of p53 by Mdmx
1YCR	B	TRP	23	136.7	64.7	-5.9	COMPLEX (ONCOGENE PROTEIN/PEPTIDE	MDM2 BOUND TO THE TRANSACTIVATION DOMAIN OF P53
3DAB	B	TRP	23	135.5	64.2	-5.8	CELL CYCLE	Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain
1YCQ	B	TRP	23	135.5	64.2	-5.9	COMPLEX (ONCOGENE PROTEIN/PEPTIDE	XENOPUS LAEVIS MDM2 BOUND TO THE TRANSACTIVATION DOMAIN OF HUMAN P53
3G03	B	TRP	23	134.0	63.4	-5.7	CELL CYCLE	Structure of human MDM2 in complex with high affinity peptide
1MR1	C	TRP	274	133.7	63.3	-8.0	SIGNALING PROTEIN	Crystal Structure of a Smad4-Ski Complex
3DAC	B	TRP	23	127.8	60.4	-5.7	CELL CYCLE	Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain
2Z5T	B	TRP	23	126.1	59.8	-5.5	CELL CYCLE	Molecular basis for the inhibition of p53 by Mdmx
1JU5	C	TRP	99	84.5	40.0	-5.0	PROTEIN BINDING/TRANSFE	Ternary complex of an Crk SH2 domain, Crk-derived phophopeptide, and Abl SH3 domain by NMR spectroscopy

Figure 2. Typical outcome of search of PPIs in PDB. Example displays entries with a Trp anchor residue contributing less than (FastContact) −5 kcal/mol towards the binding free energy and associated with the keyword 'oncogene'. Search resulted in 19 Trp residues from 15 PDBs.
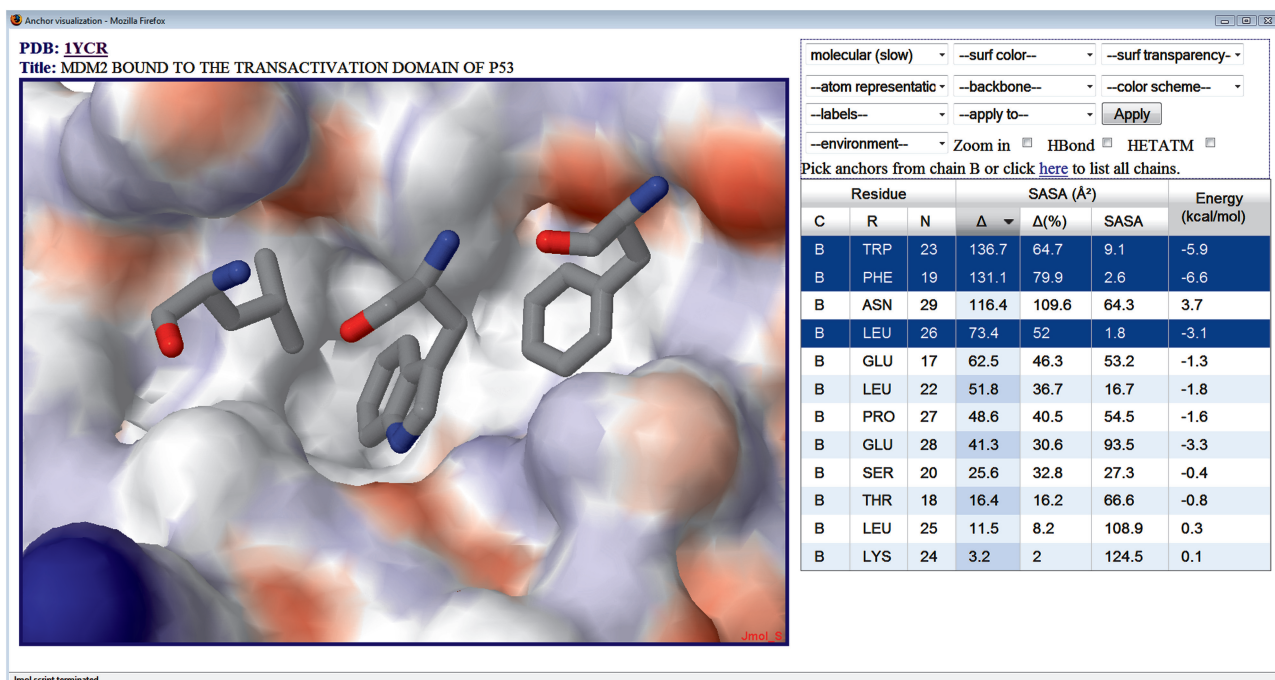
Figure 3. Visualization tool showing anchor residues Phe19, Trp23 and Leu26 of p53 from PDB 1YCR (PPI: p53-MDM2).

largest ΔSASA and lowest (i.e. favorable) predicted binding interaction energy. Note that the table of interacting residues also show the C-terminal ASN29 as having a large value of ΔSASA and unfavorable predicted energy (+3.7 kcal/mol) due to the extra COO^- group. Thus, as is, ASN29 is not a good group to target for drug design. On the other hand, the three selected residues (PHE19, TRP23 and LEU26) are indeed hotspot anchors that have been exploited on the design of compounds that bind to MDM2 (29).

Anchors versus hot-spots

It is interesting to note that although 'anchors' are often 'hot-spots', a simple screening of the ANCHOR database using the ΔSASA and Energy options shows that there are many residues with almost no buried area but strong interaction energies ('hot-spots'), as well as residues with large buried areas and little or no interaction energies ('anchors'). Although energy estimates are bound to include some false predictions, our visualization tool clearly identifies plenty of potential hot-spots with little or no value as primary target sites due to their superficial contact (often forming hydrogen bonds). The opposite is not necessarily true for deeply buried residues, since rational modifications might be able to improve the chemical affinity towards the anchor cavity.

Anchor-based drug design

Because anchor side chains play a crucial role in molecular recognition by burying large amounts of solvent-accessible surface area into the acceptor protein (9,10), side-chain analogs of anchor residues are ideal fragments to incorporate in structure-based design of compounds that might interfere with PPIs. Moreover, since there is good evidence

that in many cases the anchoring grooves are relatively unchanged upon complexation, they constitute a uniquely well-characterized site for docking small molecules. This means that anchor analogs incorporated into compounds have by design a great propensity to fill well-defined pockets in the acceptor protein.

Screening protein–protein interfaces for small molecule intervention

By using deeply buried anchor side chains as a surrogate for druggable binding sites, ANCHOR database and associated query engine can be used to screen for PPIs suitable for small molecule intervention. For example, Figure 2 shows the outcome of a search for Trp anchors with FastContact binding energy less than −5.0 kcal/mol from PPIs entries associated with the keyword 'oncogene'. The search results in 19 Trp residues from 15 PDB entries. Out of these 15 entries, 14 entries are anti-cancer PPI targets, namely: eight related structures of MDM2 and MDMX bound to p53 and short peptides; five related structures of CDK4 bound to cyclin D1 and cyclin D3; and one structure of Smad4 bound to Ski. This example highlights the potential of ANCHOR to finding relevant PPI targets for small molecule intervention.

Conflict of interest statement. None declared.

REFERENCES

1. Berg,T. (2003) Modulation of protein-protein interactions with small organic molecules. *Angew Chem. Int. Ed. Engl.*, **42**, 2462–2481.
2. Arkin,M.R. and Wells,J.A. (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.*, **3**, 301–317.
3. Fry,D.C. (2006) Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers*, **84**, 535–552.
4. Wells,J.A. and McClendon,C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **450**, 1001–1009.
5. Doemling,A. (2008) Small molecular weight protein-protein interaction antagonists: an insurmountable challenge? *Curr. Opin. Chem. Biol.*, **12**, 281–291.
6. Arkin,M. (2005) Protein-protein interactions and cancer: small molecules going in for the kill. *Curr. Opin. Chem. Biol.*, **9**, 317–324.
7. Fry,D.C. and Vassilev,L.T. (2005) Targeting protein-protein interactions for cancer therapy. *J. Mol. Med.*, **83**, 955–963.
8. White,A.W., Westwell,A.D. and Brahemi,G. (2008) Protein-protein interactions as targets for small-molecule therapeutics in cancer. *Expert Rev. Mol. Med.*, **10**, e8.
9. Kimura,S.R., Brower,R.C., Vajda,S. and Camacho,C.J. (2001) Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.*, **80**, 635–642.
10. Rajamani,D., Thiel,S., Vajda,S. and Camacho,C.J. (2004) Anchor residues in protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.
11. Clackson,T. and Wells,J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–386.
12. Bogan,A.A. and Thorn,K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1–9.
13. Thorn,K.S. and Bogan,A.A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17**, 284–285.
14. Kortemme,T., Kim,D.E. and Baker,D. (2004) Computational alanine scanning of protein-protein interfaces. *Sci. STKE*, **2004**, pl2.
15. Ofran,Y. and Rost,B. (2007) Protein-protein interaction hotspots carved into sequences. *PLoS Comput. Biol.*, **3**, e119.
16. Grosdidier,S. and Fernandez-Recio,J. (2008) Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics*, **9**, 447.
17. Lise,S., Archambeau,C., Pontil,M. and Jones,D.T. (2009) Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics*, **10**, 365.
18. Tuncbag,N., Gursoy,A. and Keskin,O. (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, **25**, 1513–1520.
19. Camacho,C.J. and Zhang,C. (2005) FastContact: rapid estimate of contact and binding free energies. *Bioinformatics*, **21**, 2534–2536.
20. Camacho,C.J., Ma,H. and Champ,P.C. (2006) Scoring a diverse set of high-quality docked conformations: a metascore based on electrostatic and desolvation interactions. *Proteins*, **63**, 868–877.
21. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
22. Tuncbag,N., Kar,G., Keskin,O., Gursoy,A. and Nussinov,R. (2009) A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief. Bioinform.*, **10**, 217–232.
23. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamic calculations. *J. Comput. Chem.*, **4**, 187–217.
24. Zhang,C., Vasmatzis,G., Cornette,J.L. and DeLisi,C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, **267**, 707–726.
25. Camacho,C.J. (2005) Modeling side-chains using molecular dynamics improve recognition of binding region in CAPRI targets. *Proteins*, **60**, 245–251.
26. Comeau,S.R., Vajda,S. and Camacho,C.J. (2005) Performance of the first protein docking server ClusPro in CAPRI rounds 3-5. *Proteins*, **60**, 239–244.
27. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
28. Xu,Q., Canutescu,A.A., Wang,G., Shapovalov,M., Obradovic,Z. and Dunbrack,R.L. Jr (2008) Statistical analysis of interface similarity in crystals of homologous proteins. *J. Mol. Biol.*, **381**, 487–507.
29. Fuller,J.C., Burgoyne,N.J. and Jackson,R.M. (2009) Predicting druggable binding sites at the protein-protein interface. *Drug Discov. Today*, **14**, 155–161.